United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

NASS Staff Report
Number SRB-87-02

December 1987

A Production Forecasting Model for Corn

Thomas R. Birkett

A PRODUCTION FORECASTING MODEL FOR CORN. By Thomas R. Birkett; National Agricultural Statistics Service; U.S. Department of Agriculture; Washington, D.C. 20250; November 1987, Staff Report No. SRB-87-02.

ABSTRACT

This report introduces a regression model for corn production that uses summarized data as input and allows for interaction between ear counts and ear lengths. In a comparison of forecasting accuracy covering the period 1980-1985, this model outperformed by a wide margin the current objective yield models in August and September and equaled the performance of the Agricultural Statistics Board in August but not in September. The method of generating estimates with this model is entirely objective. It is recommended that the current objective yield modeling system be augmented with this new model's procedures.

ACKNOWLEDGMENTS

The author wishes to thank Ben Klugh for his help in understanding the corn objective yield program and for his many suggestions that were instrumental in making this analysis successful. The author also wishes to thank Ron Steele and Fred Vogel for their suggestions in the review process.

CONTENTS

SUMMARY

This report presents a regression model for corn that produces a direct forecast of corn production for the 10 objective yield States. The model is based upon state level estimates of acres for harvest, plant and ear counts, and ear sizes. Using the season final Agricultural Statistics Board (ASB) as the final value, a comparison of forecasting accuracy was done. The ASB performance, the proposed new corn forecasting model and the current objective yield models were compared for August and September corn production forecasts. The comparison showed the proposed corn model outperformed the current objective yield models by a wide margin in August and September and outperformed the ASB in August but not September.

The current objective yield procedure uses separate sample level regression models for final ear counts and ear weights. The approach presented in this paper differs from the current procedure in the following ways: (1) the sample level data are summarized to the 10-State level (State level for the State models) before being entered in the model, and (2), the ear count by ear size interaction term is included in one model, instead of having a separate model for each variable. The range of the model includes normal yielding years as well as the drought years 1980 and 1983 and the high yielding 1985. The method of generating estimates with this model is entirely objective. A 90 percent prediction interval for August would be approximately plus or minus 10 percent.

Many relationships are examined with the discovery, at the regional level, that two variables do consistently well in August and one variable outperforms all others in September. These variables are the product of an acreage, count and size variable. The model form is $Y = bX$, so it is a no intercept model and has only one parameter, b. The variables that do well in August are X = *(total acres)* × *(number of ears with kernels per acre)* × *(average kernel row length per ear)*, and X = *(total acres)* × *(number of stalks with ears per acre)* × *(average kernel row length per ear)*. In September the superior variable is X = *(total acres)* × *(number of ears and ear shoots per acre)* × *(average kernel row length per ear)*. State models are also developed, and the regional model is used to prorate State estimates to the predicted value of the regional model.

It is recommended the proposed forecast model be used to augment the current NASS objective yield modeling system.

A Production Forecasting Model for Corn

Thomas R. Birkett

INTRODUCTION

This paper develops a regression model for forecasting corn production that shows accurate predictive capabilities across a broad spectrum of years.

The model was developed using the corn objective yield data for 1980-1985. The development is analogous to a technique used for forecasting yields of fruit and nut crops the author developed while working in the California State Statistical Office. The functional form of the independent variable is found in the work of Fecso (1975).

The statistical philosophies behind the existing model and the proposed model are summarized below:

Existing objective yield models:

(1) The models are built with sample-level data, so an individual sample can have a large influence on the parameter estimates. These are straight-line models and their parameter estimation procedures can be unstable in the presence of the numerous extreme values that exist in the sample-level data. An automated procedure has been developed to identify and remove these extreme values, restoring some stability. However, by deleting observations, the properties of the estimated parameters are no longer necessarily optimal. Existence of the outliers also provides evidence that a straight-line model does not adequately describe the underlying relationship between the dependent and independent variables over the entire measurement space.

(2) There are separate models for final number of ears and ear weight. The lack of model fit in each model may be compounded when the results are combined.

(3) Interaction between ear counts and ear sizes is not included until the predicted values from the two models are combined.

Proposed Objective Yield Model:

(1) Unbiased estimates of the 10 State level mean number of ears per square foot and mean length per ear are constructed from the sample level data and the State acreages. Due to the large sample size, independent samples, and the central limit theorem the approximate distribution of these estimators is known. Any individual sample can have only a small effect upon these estimated parameters. In particular the variance is small because of the large sample size.
(2) One model that includes the ear count by ear size interaction is used. The independent variables will be those calculated in (1) and consequently very stable.
(3) This model is based on the same input as existing procedures, but should much more closely approximate the underlying relationship between the dependent and independent variables.

The remainder of this paper has the following format. First, the basic algebraic formulas for the possible versions of the independent variable X are presented. The survey design allows the construction of 64 possible combinations of *(acreage variable)* × *(count variable)* × *(size variable)*. Unbiased estimators for number of counts per acre and average size per ear are defined as part of the construction of X. Next, in the Methods section, a way of selecting the best version of X is presented, along with a method for combining forecasts from more than one X. The Results section presents the data set and discusses the performance of the superior X's, the Agricultural Statistics Board, and the current objective yield models in predicting the final Agricultural Statistics Board production. The performance of each is compared and conclusions are drawn based on these results.

THE MODEL

The model has the form $Y = bX$, and is consequently a one variable no-intercept model. (An hypotheses test for a zero intercept has a p-value of 0.8). When the variable X contains acreage information, Y is production. When it does not, Y represents yield. In the development in this paper, acreage information is included as part of X, so Y represents production. The results are almost as good if X does not include acreage and Y is yield. When X does not

include acres, the predicted yield would be multiplied by the acreage estimate to get production. The methodology can be implemented either way.

The predictor variable X for production is constructed from variables collected in the objective yield survey, combined with acreage information from the June Enumerative Survey (JES). The variable X is a function of all information from all the samples in all States for each month. The underlying relationship between this variable and final corn production will be modeled by a straight line through the origin. There consequently is only one parameter to estimate and the model can be fit with as little as one year of historical data. The zero intercept also sharply reduces the variance of the slope parameter estimate.

The JES and the corn objective yield survey provide the following variables: (1) State acres — the JES Harvest to Planted *(jes/hp)* and the JES Direct Expansion *(jes/de)*; (2) sample level counts per square foot — number of stalks *(s)*, number of stalks with ears *(se)*, number of ears and silked ear shoots *(eaes)*, and number of ears with kernels *(ek)*; and (3) sample level ear sizes — mean kernel row length *(krl)* and mean length over husk *(loh)*, each from a subsample of ears.

The count variables are transformed to a square foot basis because the area of each sample varies (due to the differing row widths). For this transformation, it is assumed that the area of each sample is equal to the area of the rectangle formed by the 15-foot measurement and the eight row width measurement divided by two. The two size variables are also weighted by the counts per square foot at the sample level to obtain unbiased estimates of the mean size per ear. Preliminary analysis led to limiting the analysis to samples with maturity 3 or greater. (The maturity coding system for the samples is 1-2, before ears are present; 3-6, ears present; and 7, harvest. Almost all samples have maturity greater than 2 after August). In the formulas that follow, all sample counts have been converted to counts per square foot.

The construction of the independent variable in the model, $X_{jk_1k_2l}$, is a series of straight and weighted means. The dependent variable, Y, is Board production in billions of bushels. The series of weightings starts with the sample-level data and continues in stages to the 10-State regional level. The sample is considered random inside of each State. And, because of the transformation of the count data,

there is a random sample of square feet within each State. The number of ears per square foot varies, so the size variables are weighted by the number of ears they represent. The State acreages also vary.

Unbiased estimates of the individual State and 10-State regional mean counts per square foot and mean sizes per ear are derived separately. (Estimated values of these parameters for 1980-1985 for the 10-State region and for Iowa are plotted in figures 1 and 2 of Results.) These means are multiplied together at the State and regional levels. In the following, the references to the two acreage variables, four count variables, and two size variables are to those variables already described as making up the input data. Starting with the sample-level data, the construction of $X_{jk_1k_2l}$ has the following sequence.

Let I denote the set of all States in the survey in any given month, and let I^* denote the set of States in the survey in any given month that has at least one completed sample with size data (of maturity 3 or greater). In the following, the subscript $i \in I$ will always refer to the States, the subscript $j \in \{1,2\}$ will represent the acreage variable, k_1 and $k_2 \in \{1,2,3,4\}$ will refer to the count variables, and $l \in \{1,2\}$ will always represent the size variable.

The independent variable $X_{jk_1k_2l}$ has the functional form *(regional acreage estimate)* × *(regional mean count per square foot estimate)* × *(regional mean size per ear estimate)*:

$$X_{jk_1k_2l} = A_j \, \overline{C}_{jk_1} \, \overline{S}_{jk_1k_2l} \tag{1}$$

where A_j is the sum of the State acreage estimates for acreage variable j and is defined as:

$$A_j = \sum_{i \in I} a_{ij} \tag{2}$$

where a_{ij} = the acreage for State i, acreage variable j.

$\overline{C}_{jk_1}$ is the weighted average of the State mean counts per square foot, weighted by the State acreages, and is defined as

$$\overline{C}_{jk_1} = \frac{\sum_{i \in I^*} a_{ij} \, \overline{c}_{ik_1}}{\sum_{i \in I^*} a_{ij}} \tag{3}$$

where $\overline{c}_{ik_1}$ = the mean count per square foot for count variable k_1 in

State i.

$$\overline{c}_{ik_1} = \frac{1}{n_i} \sum_{m=1}^{n_i} c_{ik_1m} \tag{4}$$

where c_{ik_1m} is the count per square foot for count variable k_1, State i, sample m, and n_i is the sample size in State i for samples with maturity $\in \{3,4,5,6\}$.

$\overline{S}_{jk_1k_2l}$ is the regional mean size per ear, weighted from the sample level to the State level by count variable k_2, and weighted from the State level to the regional level by *(acreage variable j)* $\times$ *(count variable* k_1).

$$\overline{S}_{jk_1k_2l} = \frac{\sum_{i \in I^*} a_{ij}\, \overline{c}_{ik_1}\, \overline{s}_{ik_2l}}{\sum_{i \in I^*} a_{ij}\, \overline{c}_{ik_1}} \tag{5}$$

where $\overline{c}_{ik_1}$ is defined above, $\overline{s}_{ik_2l}$ is the mean size per ear for size variable l in State i, using as sample level weights count variable k_2.

$$\overline{s}_{ik_2l} = \frac{\sum_{m=1}^{n_i} c_{ik_2m}\, s_{ilm}}{\sum_{m=1}^{n_i} c_{ik_2m}} \tag{6}$$

$s_{ilm} =$ the size per ear for size variable l, State i, sample m, c_{ik_2m} is the count per square foot for count variable k_2, State i, sample m, $n_i =$ number of samples in State i with maturity $\in \{3,4,5,6\}$.

METHODS

The computer was used to calculate the values of the 64 $X_{jk_1k_2l}$ defined in (1). Values for State models were also calculated.

There were, again, two acreage variables, four count variables, and four different weighted versions of each of two ear size variables. This results in 64 possible variable combinations of the form *acreage* × *count* × *size*, the general form for X.

Modeling all 64 combinations is a systematic way of maximizing the information gained from the survey. The fact that the designers of the survey chose to include certain variables is sufficient reason to model them. After examining analyses such as these the designers may decide that certain variables need no longer be collected. The reason for modeling all four weighted versions of the two size variables is, a priori, it is possible that one weighting count variable may outperform the others. Later in this paper it will be examined whether some combining of variables can take place, thus reducing the overall number.

Table 1 portrays the factorial process used to construct the 64 *acreage* × *count* × *size* variables. This table is a pictorial representation of the algebraic derivations in the previous section. From each variable the model $Y = bX$ is fit with b estimated with least squares (and Y is Board production).

Table 1 — Construction of 64 possible independent variables for the model by multiplying together one acreage, one count, and one size variable, and eight additional independent variables obtained by multiplying one acreage and one count variable. Columns are (j):(k_1):(l) (k_2).

Acreage	Count	Size (with sample-level weight)
		kernel row length (stalks)
		kernel row length (stalks with ears)
	stalks	kernel row length (ears and ear shoots)
jes/hp	stalks with ears	kernel row length (ears with kernels)
jes/de	ears and ear shoots	length over husk (stalks)
	ears with kernels	length over husk (stalks with ears)
		length over husk (ears and ear shoots)
		length over husk (ears with kernels)
	stalks	
jes/hp	stalks with ears	
jes/de	ears and ear shoots	
	ears with kernels	

As an example X_{1111} is *(jes/hp)* × *(stalks)* × *(kernel row length weighted by stalks at the sample level)*. X_{2442} is *(jes/de)* × *(ears with kernels)* × *(length over husk weighted by ears with kernels at the sample level)*. Also included in the table are eight *acreage* × *count* variables whose purpose is explained below.

Because some States do not have any completed samples with maturity 3 or greater in August, it is not always possible to calculate State values for the 64 $X_{jk_1k_2l}$ in the first month of the survey. This is because if there are no samples with ears then there is no estimate of the average size per ear (the $\bar{s}_{ik_2l}$). To fill this void, eight additional variables were created, corresponding to the two acreage variables multiplied by the four count variables. Values for these eight variables will always be available, and their addition results in a total of 72 possible independent variables to choose from for each month of the survey before harvest. Preliminary analysis resulted in limiting the first 64 model variables to samples with maturities of 3 or greater. The last eight, which lack a size component, include all samples.

In August, the number of samples with size data is limited, and, as stated earlier, some States won't have any samples with maturity 3 or greater. For these States, the model is limited to the eight *acreage* × *count* variables. To calculate the value of the 64 $X_{jk_1k_2l}$ for the 10-State regional model in August, the method calculates the mean counts and mean sizes for the subregion of States that do have size data (the set I^*), and then applies these means to the 10-State region. Because almost all samples have reached maturity 3 by September, the variable X for September is based on a large amount of data.

COMPOSITE FORECAST

At this point what is available from this system are 72 one-variable no-intercept regression models. The question is, what variable should be used to make the forecast in the current year, or how can the forecasts from several models best be combined? The reason for not considering a regression model with more than one independent variable in it at this point is because the number of years in the model precludes this. With $n = 6$ it is better to limit the number of parameters in each regression to one. This restriction can be lifted as more years of data become available. Consequently, to use the information from more than one variable we will use a average of the predicted values from each one variable

model. The method of combining the forecasts is as follows. (See Houseman).

From a preliminary analysis it was determined that the combining of forecasts would work better if the set of 64 variables with a size component were combined to 16. This combining involved averaging across the k_2 subscript of each set of four $X_{jk_1k_2l}$ with the same j, k_1, and l subscripts.

$$X_{jk_1l} = \frac{1}{4} \sum_{k_2=1}^{4} X_{jk_1k_2l}. \tag{8}$$

k_2 represents the sample level weight on the size variable. The interpretation of this procedure is that the sample level weight on the size variables now becomes the average of the four count variables. After this is done there are 16 variables with a size component and 8 variables without one, for a total of 24 variables. The 16 now represent the 2 acreage by 4 count by 2 size factorial generation of all possibilities. The 8 represent the 2 acreage by 4 count generation of possibilities. At this point the forecasting system has 24 one variable regression models generating 24 $\hat{Y}$'s.

$$(\hat{Y}_1,\hat{Y}_2,.....,\hat{Y}_{24}) \tag{9}$$

The composite model is the average of the forecasted values from a subset of the 24 available forecasts. (In this case of composite estimation it was found that equal weights were optimal, once the members of the composite model were selected). This subset was chosen to be the subset that had minimum estimated variance in predicting final Board production. (It was not possible to look at all possible subsets, so a forward selection type procedure was utilized to select the best subset). Stated more explicitly, the subset of the 24 $\hat{Y}_i$'s whose mean has minimum estimated variance is the composite model.

$$\hat{Y} = \frac{1}{k} \sum_{i=n_1}^{n_k} \hat{Y}_i \tag{10}$$

where $(\hat{Y}_{n_1},...,\hat{Y}_{n_k})$ are the subset members whose mean has minimum variance over all possible subsets.

To estimate the stability and error rate of this procedure a jacknife type of analysis was devised. For each combination of five years taken from the six years 1980-1985, of which there are 6 (combinations), the minimum variance criterion identified the subset of variables to make up the composite model. The mean predicted value

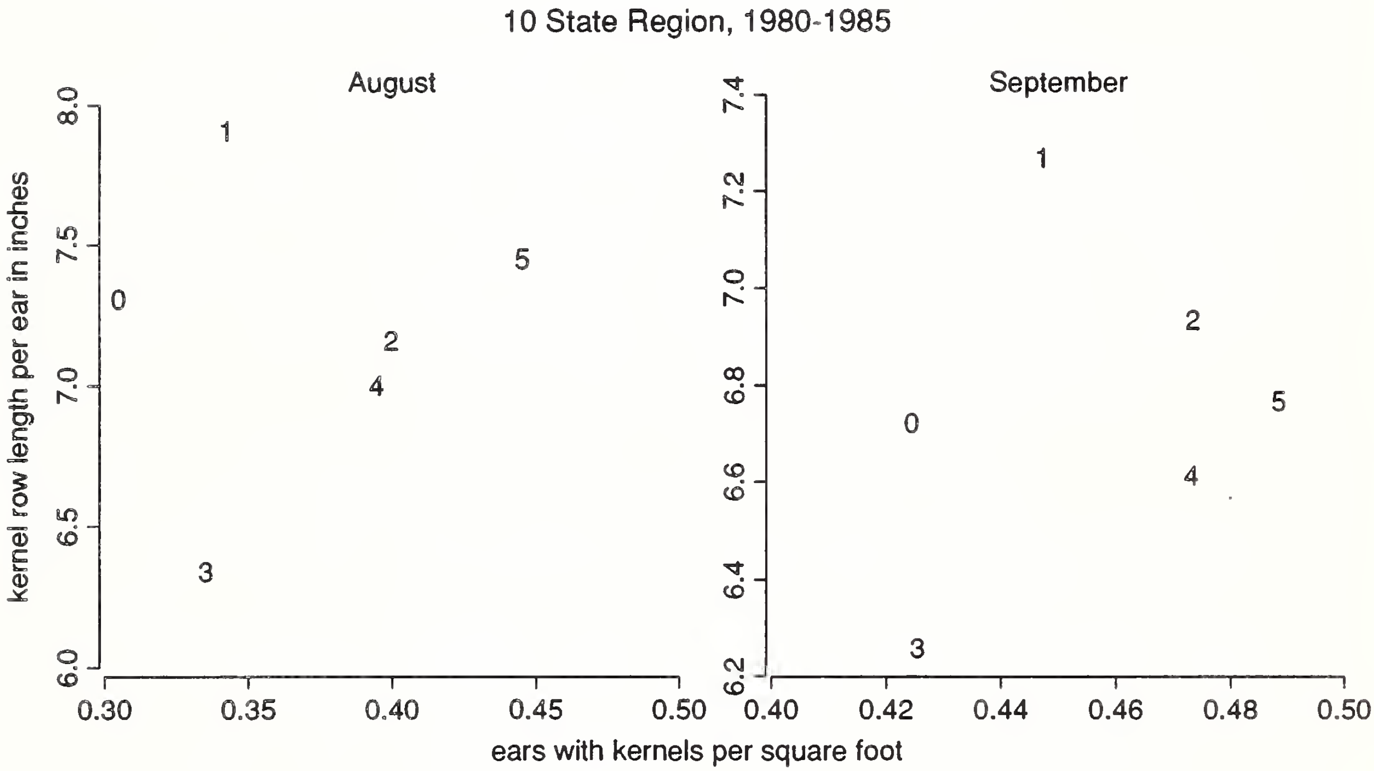
of this subset was determined, along with the difference between it and the actual value of Y. In this method the predicted value and the actual value are functions of different years and therefore independent. Jacknifing is a way of guarding against the possibility that so many combinations were looked at that a random good fit would occur, even though it would not hold in reality. Detailed results from this jacknife appear in the next section. The variables in the composite model each year and the prediction errors are presented so one can observe the stability and accuracy of this procedure. This method is compared to the record of the Agricultural Statistics Board (ASB) over the same period.

RESULTS

Observed values of the ears with kernels per square foot and the kernel row length per ear are shown in figure 1 and figure 2.

Figure 1 presents plots for the 10-state region and figure 2 for Iowa covering 1980-1985. The August values are from only the samples with maturity greater than or equal to 3 (ears present).

Figure 1 — Observed values of ears with kernels per square foot and average kernel row length per ear, 1980-1985, for the 10 state region. The symbol represents the year.



These plots illustrate why the model works. High yielding years appear in the upper right hand corner, because they have both large sizes and high counts. The lowest yielding year is in the lower left hand corner, where the years with low counts and low sizes are located. Also the general orientation does not change when going from August to September, which supports the validity of the early August size data, as it is modeled here.

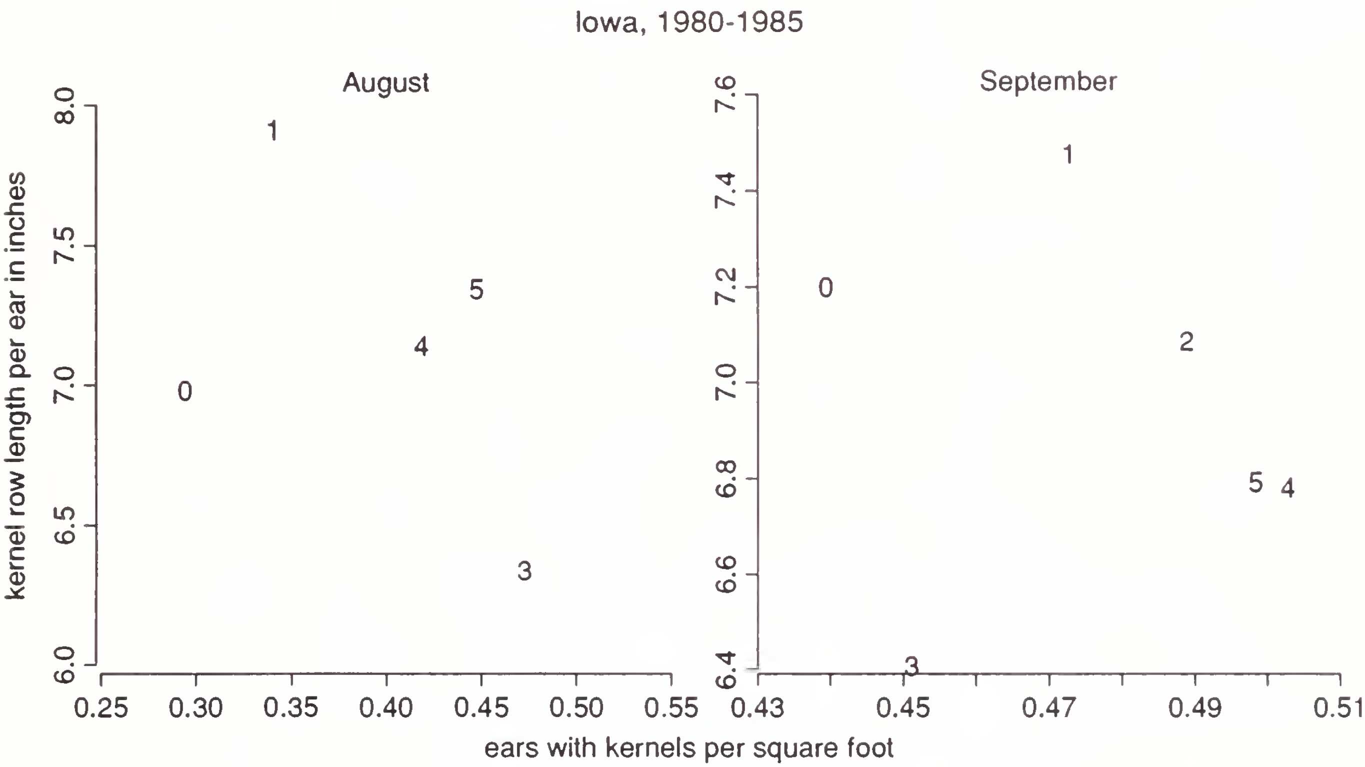
The observed data indicates that in August 1983 the survey picked up early indications of low counts and very small sizes. 1983 was a drought year in the Midwest. The survey also picked up early signs of low counts in 1980, another year with below average

rainfall.

1981 had normal counts and very large sizes. 1982 and 1984 are characterized by normal sizes and high numbers of ears. As mentioned, 1980 had very low numbers and normal sizes, while 1983 had the lowest yields of the period with somewhat below normal counts and small ears.

In August 1985 the counts of ears with kernels were the highest of the period, and the ear sizes were above average. In September of that year, the counts remained at record high levels, but the sizes were in the average range. Heavy rains fell in the corn belt between the August and September surveys that year, slowing down the growth in ear size.

Figure 2 — Observed values of ears with kernels per square foot and average kernel row length per ear, 1980-1985, for Iowa. The symbol represents the year.



The pattern for Iowa follows the regional pattern. In 1982 in Iowa there were no samples with maturity 3 in August.

Figure 3 presents the plot for the best single variable August model, *ears with kernels* × *kernel row length* (*ek* × *krl*). Notice the relationship is very linear through the origin. Later it will be shown that this variable in conjunction with one other variable form the best composite model for August.

Figure 3 — August model with *ears with kernels* × *kernel row length* (*ek* × *rl*). The symbol represents the year.

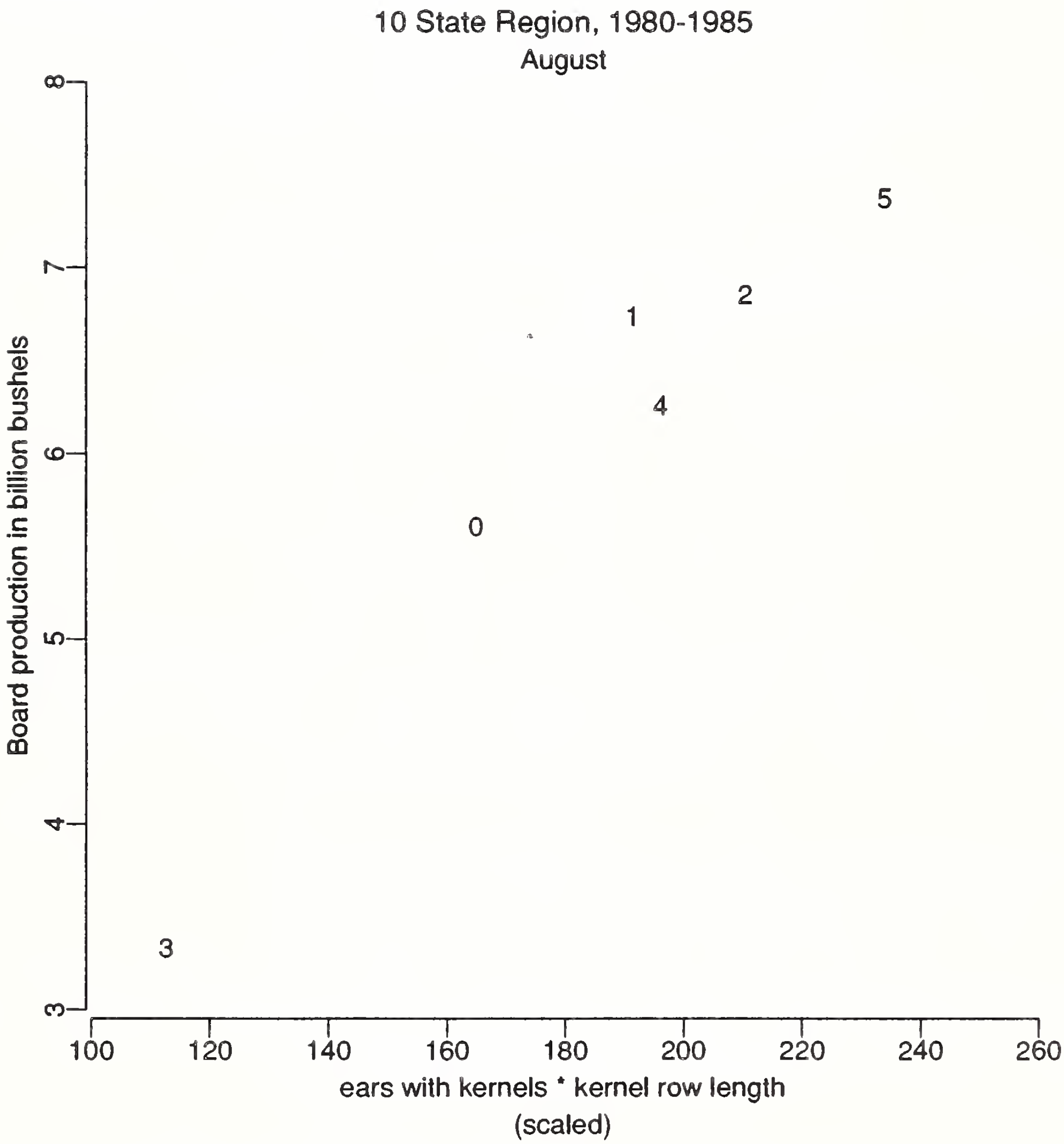


Figure 4 exhibits a plot of the best September variable, *ears and ear shoots × kernel row length* (*eaes × krl*). The variation around the fitted line is small, with the largest residual at 1985.

Figure 4 — The September model with *ears and ear shoots × kernel row length* (*eaes × krl*). The symbol represents the year.

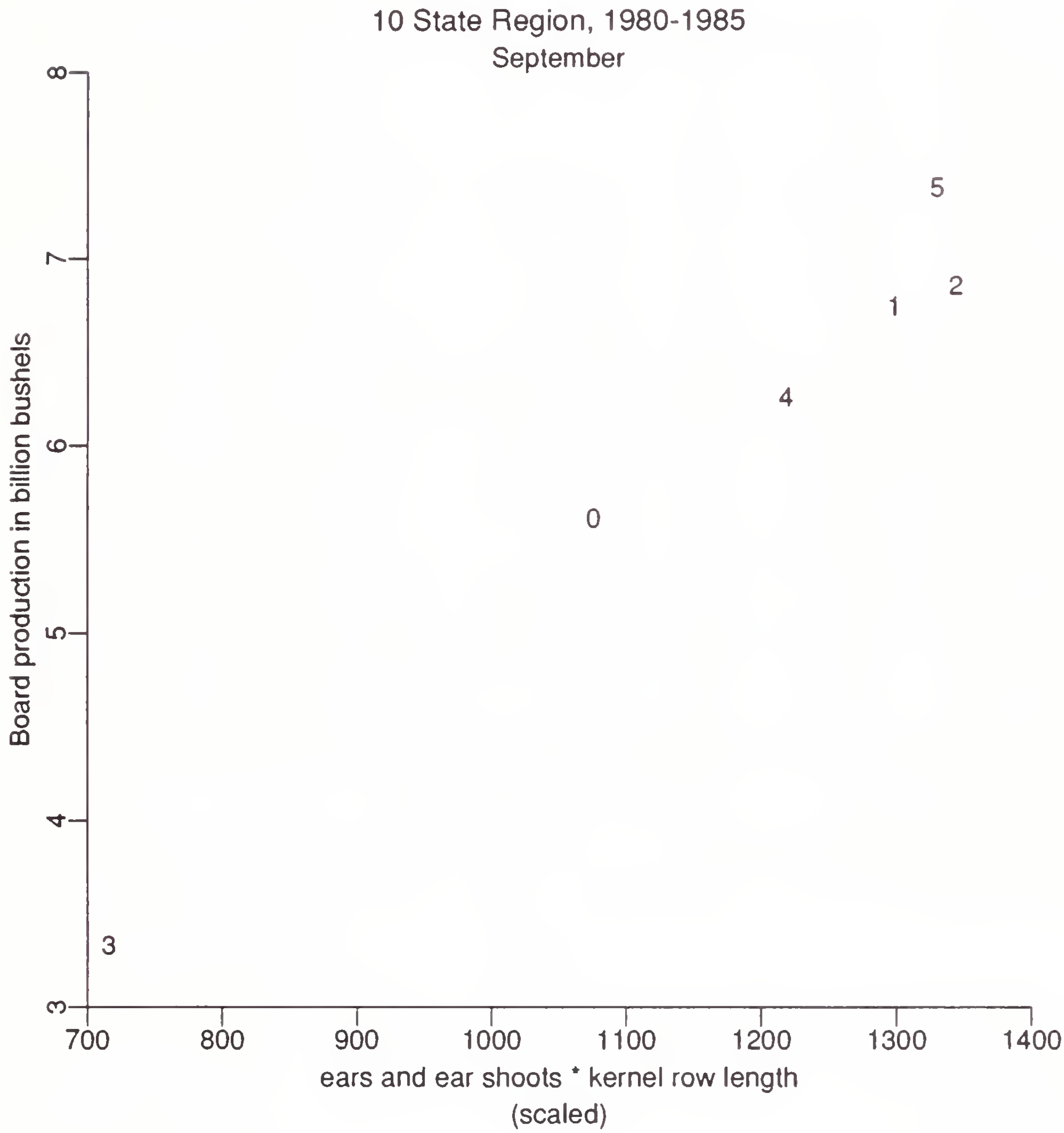


Table 2 exhibits the performance of the composite model.

Table 2 — 10 State jacknife results, 1980-85.

Month	Year	Variables in Composite Model 1/	Model Predicted Value	Board Predicted Value	Board Final	Model Error	Board Error
			billion bushels			%	
August	80	*ek* × *krl* *se* × *krl*	5.74	5.60	5.61	2.5	0.2
	81	*ek* × *krl* *se* × *krl*	6.40	6.40	6.74	4.3	5.0
	82	*ek* × *krl* *se* × *krl*	6.48	6.96	6.85	5.4	1.6
	83	*ek* × *krl* *se* × *krl*	3.68	4.27	3.34	10.2	27.8
	84	*ek* × *krl* *se* × *krl*	6.24	6.32	6.26	0.3	1.0
	85	*ek* × *krl* *se* × *krl*	7.51	6.84	7.38	1.8	7.3
September	80	*eaes* × *krl*	5.58	5.54	5.61	0.5	1.2
	81	*eaes* × *krl*	6.77	6.59	6.74	0.4	2.2
	82	*eaes* × *krl*	7.04	6.96	6.85	2.8	1.6
	83	*eaes* × *krl*	3.75	3.53	3.34	12.3	5.7
	84	*eaes* × *krl*	6.45	6.17	6.26	3.0	1.4
	85	*eaes* × *krl*	6.81	7.02	7.38	7.7	4.9

1/ Variable abbreviations *s*-stalks, *se*-stalks with ears, *eaes*-ears and ear shoots, *ek*-ears with kernels, *krl*-kernel row length, *loh*—length over husk.

In August the number of variables in the composite is two, *ears with kernels* × *kernel row length* (*ek* × *krl*), and *stalks with ears* × *kernel row length* (*se* × *krl*).

In September the number of variables in the composite model is one, ears and ear shoots by kernel row length (*eaes* × *krl*).

Since the August and September models are constant, there is evidence that the system contains an inherent stability.

This system will now be compared to the Agricultural Statistics Board (ASB) and the current objective yield models.

Table 3 — Average prediction errors of this objective yield model, the ASB, and the current objective yield models, based on 1980-85.

Month	State	Average percent prediction error		
		Model	ASB	Current OY Models
August	Illinois	10.3	10.3	20.1
	Indiana	10.0	9.0	16.5
	Iowa	7.6	9.4	12.6
	Michigan	8.2	5.8	6.3
	Minnesota	7.4	9.3	9.9
	Missouri	16.5	16.8	37.6
	Nebraska	7.5	7.4	11.9
	Ohio	13.5	12.5	14.5
	South Dakota	22.2	9.9	21.5
	Wisconsin	2.2	5.6	6.1
State wghtd. avg. 1/		9.3	9.4	11.4
September	Illinois	7.1	2.6	9.2
	Indiana	6.6	2.3	11.5
	Iowa	8.7	3.0	7.9
	Michigan	8.2	3.6	4.5
	Minnesota	7.0	6.9	8.1
	Missouri	14.0	8.3	17.8
	Nebraska	4.6	5.9	10.8
	Ohio	8.4	7.3	14.1
	South Dakota	12.5	12.6	15.5
	Wisconsin	1.1	3.0	3.2
State wghtd. avg. 1/		7.3	4.4	8.0

1/ The state errors are weighted by state production.

Table 3 presents the results for the composite model developed in the last section. Forecasts were made for the 10 state region and for each state individually. In August each state model except for Missouri was number of stalks (s). The Missouri model was ears and ear shoots (*eaes*). None of the August State models utilized size data. In September six state models consisted of *ears and ear shoots* × *kernel row length* (*eaes* × *krl*), and the other four were *ears with kernels* × *length over husk* (*ek* × *loh*). In each month the state forecasts were then prorated to sum to the predicted value from the 10 state model. The objective yield indication is adjusted for bias according to the current Board charting system.

In August the model was equal with the ASB (9.3 to 9.4 percent average error), and 18 percent more accurate than the objective

yield models currently used (9.3 to 11.4).

In September the ASB improved considerably to an average error of only 4.4 percent, while the model improved to 7.3 percent average error. The model outperformed the current objective yield models by 9 percent (7.3 to 8 percent).

An examination of the errors in Table 2 reveals that August 1983 was the year where this model significantly outperformed the ASB. In that drought year the model recognized the small ear sizes and low counts before the Board. In more normal yielding years the errors of the Board were on average smaller.

DISCUSSION

The two primary criticisms of this approach are that it uses a limited number of years and that it does not outperform the ASB (although they are on average equal in August). These concerns will now be addressed.

The philosophy behind limiting the number of years is based on making the model adaptable to changing conditions. The correctness of the model specification depends on the relationship being a straight line through the origin. Normally this relationship cannot be expected to hold over long series of years, due to changes in underlying conditions that induce a time trend. The mechanism for adjusting for this time trend is to delete years from the beginning of the series until the nonlinearity is removed. For this data set the relationship is linear beginning in 1980. The model is deliberately specified to have only one slope parameter and a small variance around the fitted line, so that very small n is sufficient. Also each observation represents an entire survey of data, so the amount of information going into the model is equal to the current models. It could be said that this model is able to use even more information because it doesn't require that outliers in the raw data be modified of deleted.

The second criticism is that the model does not do as well as the ASB in September. One aspect that is disadvantageous to the model's historical performance is that this performance is measured at the State level. The Board uses a balance sheet approach to force the state sums to a prespecified total, and their allocating procedures are subjective. This can cause some problems for the state level models, which remain objective functions of acreage, counts and sizes. Also, the measure of forecasting accuracy for the September ASB depends on the final ASB, and the two may not be

entirely independent.

CONCLUSION

Based on these results this approach is a better way to model the objective yield data than that currently used. This paper has presented the exact form of the model and how to form the estimates of counts per unit area and average sizes. These estimates are seen to be unbiased, and the form of the model satisfies underlying regression assumptions very well. Confidence intervals on predicted values for production are an immediate consequence of this formulation. The work of Warren and Cook on developing an early season estimate of final ear weight using weather data also fits into this framework. The current estimate of size per ear is interchangeable with their estimator. The carefully defined structure of the independent variable as being the product of three separately estimated components allows this flexibility. Because of these characteristics and others already presented, this approach can be said to be well documented with superior results to the current objective yield modeling system.

Also as more years become available we may be able to add another variable to the regression (use a two-variable no-intercept model) and not be tied to composite estimation.

RECOMMENDATIONS

I recommend that the current objective yield modeling system for corn be augmented with this simple, direct, and competitive modeling system.

BIBLIOGRAPHY

[1] Allen, D.M., "Mean Square Error of Prediction as a Criterion for Selecting Variables," *Technometrics*, 13, 1971, pp. 469-475.

[2] DeGroot, Morris H., *Probability and Statistics*, Don Mills, Ontario: Addison Wesley, 1975.

[3] Fecso, Ronald S., "A Study of Walnut Production Forecasting," Staff Report, California Crop and Livestock Reporting Service, 1975.

[4] Houseman, Earl E., "Composite Estimation," Staff Report, U.S. Department of Agriculture, Statistical Reporting Service.

[5] Rao, C. Radhakrishna, *Linear Statistical Inference and Its Applications*, New York: John Wiley and Sons, 1973.